

United States Patent Application

For

**Improved Resource Allocation
in a Circuit Switched Network**

Inventors:

Vivek Gupta
Sumit Sanyal
Xiangzhong Zeng
Stephen J. Sifferman
Leah J. Fera

Prepared by:

Blakely, Sokoloff, Taylor & Zafman, LLP
12400 Wilshire Boulevard
Seventh Floor
Los Angeles, CA 90025-1026
(503) 684-6200

EL034437135US

information is to be exchanged between the ASP and a subscriber unit, another connection is set up and later torn down at the appropriate times to handle the transmission of subsequent data packets between the ASP and the subscriber unit.

However, a connection in a circuit switched network, for example, between BS
5 235 and a subscriber unit, typically requires setting up a dedicated communications channel between the SU and the BS. The channel remains allocated even when no packets are being transmitted between the BS and the SU, and is torn down only when requested to do so by either the BS or the SU. There may be significant periods of time during which the allocated channel is idle.

10 The set up and tear down of a communications channel in a circuit switched network typically consumes more time and overhead than in a packet switched network, and the channel may be underutilized, for example, by remaining idle for extended periods of time. Additionally, there may be many more subscriber units, each capable of establishing multiple simultaneous communications channels with the BS, than there are
15 communications channels capable of being simultaneously allocated.

Additionally, there is the well known concept of quality of service, or grade of service, in a packet switched network environment, in which factors such as minimum latency and bandwidth are considered. When communicating packet switched data from a subscriber unit over a circuit switched network, and then to a destination device, such
20 as ASP 205, coupled to the packet switched network, a resource allocation scheme is needed in the circuit switched network to meet or exceed the minimum quality of service for data transmissions in the packet switched network.

It is understood that well known wireless communication protocols, such as CDPD, GPRS and EDGE, employ resource allocation schemes to efficiently allocate RF channels. However, such protocols are thought to rely solely on the amount of data queued for transmission by a particular subscriber unit relative to other subscriber units.

- 5 What is needed is an allocation scheme that considers other equally or more pertinent factors.

Summary

- The present invention contemplates allocating resources in a circuit switched data
- 10 network. The process comprises receiving a request for a resource from a device coupled to the circuit switched data network and granting the resource to the requesting device if the resource is available. However, if the resource is not available, the process involves examining the instantaneous quantity of data to be transmitted by the requesting device;
- 15 the rate of change in the instantaneous quantity of data to be transmitted by the requesting device; and the time of utilization of the resource by the requesting device. The resource is granted to the requesting device based on a consideration of at least these three factors.

Brief Description of the Drawings

- The present invention is illustrated by way of example, and not necessarily by
- 20 way of limitation in the figures of the accompanying drawings in which like reference numerals refer to similar elements.

Figure 1 is a flow diagram of a process in one embodiment of the present invention.

Figure 2 is a diagram of a packet switched communications over a circuit
5 switched data network, and in which an embodiment of the present invention operates.

Detailed Description of the Invention

Reference in the specification to “one embodiment” or “an embodiment” means
that a particular feature, structure, or characteristic described in connection with the
10 embodiment is included in at least one embodiment of the invention. The appearances of
the phrase “in one embodiment” in various places in the specification are not necessarily
all referring to the same embodiment.

Some portions of the detailed description that follow are presented in terms of
algorithms and symbolic representations of operations on data within a computer
15 memory. These algorithmic descriptions and representations are the means used by those
skilled in the data processing arts to most effectively convey the substance of their work
to others skilled in the art.

An algorithm is here, and generally, conceived to be a self-consistent sequence of
steps leading to a desired result. The steps are those requiring physical manipulations of
20 physical quantities. Usually, though not necessarily, these quantities take the form of
electrical or magnetic signals capable of being stored, transferred, combined, compared,
and otherwise manipulated. It has proven convenient at times, principally for reasons of

common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like.

It should be borne in mind, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities. Unless specifically stated or otherwise apparent from the following discussion throughout the description, discussions using terms such as "processing" or "computing" or "calculating" or "determining" or "displaying" or the like, refer to the action and processes of a computer system, or similar electronic computing device, that manipulates and transforms data represented as physical (electronic) quantities within the computer system's registers and memories into other data similarly represented as physical quantities within the computer system memories or registers or other such information storage, transmission or display devices.

The invention also relates to apparatus for performing the operations herein. These apparatus may be specially constructed for the required purposes, or may comprise a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a machine-readable storage medium, such as, but not limited to, any type of magnetic or other disk storage media including floppy disks, optical storage media, CD-ROMs, and magnetic-optical disks, read-only memories (ROMs), random access memories (RAMs), EPROMs, EEPROMs, magnetic or optical cards, flash memory devices; electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.); etc. or any type of media suitable for storing electronic instructions, and each coupled to a computer system bus.

The algorithms and displays presented herein are not inherently related to any particular computer or other apparatus. Various general-purpose systems may be used with programs in accordance with the teachings herein, or it may prove convenient to construct more specialized apparatus to perform the required method steps. The required
5 structure for a variety of these systems will appear from the description below. In addition, the present invention is not described with reference to any particular programming language. It will be appreciated that a variety of programming languages may be used to implement the teachings of the invention as described herein.

In one embodiment of the invention, resources are allocated on demand. In
10 particular, resources are allocated for the transmission of packet switched data over a circuit switched network such as illustrated in figure 2. While the circuit switched network illustrated in figure 2 represents a “wireless local loop” network between a number of subscriber units and a base station/mobile switching center, it will be apparent that the present invention is likewise applicable to other circuit switched data networks,
15 for example, an Asynchronous Transfer Mode (ATM), Frame Relay, X.25 or other circuit switched data networks. Thus, the present invention may be utilized to allocate resources among devices in any circuit switched data network environment and is not limited to the wireless local loop environment comprising a base station and a number of subscriber units, as herein described as an example embodiment of the present invention.

20 In one embodiment of the present invention, the resource being allocated among subscriber units is a number of radio frequency (RF) channels. A base station controls the allocation of such RF channels to a number of subscriber units. In one embodiment, the subscriber units outnumber the RF channels. Moreover, the subscriber units may

request and be simultaneously allocated multiple RF channels. It is appreciated, however, that the inventive techniques described herein may equally be applicable to the allocation of other resources common to the transmission of data in a circuit switched network, for example, memory or buffer space, or other resources utilized and shared among a number of devices. However, for purposes of understanding the invention, the description herein will be focused on allocating communication channels in a circuit switched data network, and in particular, allocating RF communication channels in a wireless circuit switched data network environment.

With reference to figure 2, the present invention is contemplated as being utilized in a circuit switched network carrying packet switched data. A central base station (BS) has a maximum capacity of some number of RF channels operating simultaneously. An algorithm implemented in the BS allocates the RF channels to individual subscriber units, considering three factors:

- 1) the data rate of each subscriber unit;
- 2) the rate of change of the data rate of each subscriber unit; and
- 3) the amount of time for which each subscriber unit has been allocated an RF channel resource.

With respect to the first above mentioned factor, RF channels are allocated to a subscriber unit in proportion to the SU's data rate relative to the data rate of other SUs competing for the same RF channel resources. Data rate, as that phrase is used herein, refers to the rate at which data in the form of data packets is received and queued at the

SU for transmission over the circuit switched network. Additionally, as per the second above mentioned factor, the BS takes further into consideration the trend or change in the data rate for each SU. In another embodiment, the algorithm operating at the BS additionally considers the amount of time for which a subscriber unit already has been allocated a number of RF channels. In yet another embodiment, the BS considers a moving average of one or more of the above described three factors in addition to or in place of consideration of the three factors. This is particularly advantageous if there is a relative spike or peak in one or more of the values for the three factors. Thus, if the extent to which the most recent measurement of one of the factors diverges significantly from previous measurements for the factor, the most recent measurements may be discarded or at least given less of a weighting factor.

In one embodiment, the BS is capable of associating a weighting factor with one or more of the above three factors, to vary the extent to which the BS relies on any one particular factor in deciding whether to allocate an RF channel to a subscriber unit. For example, given bursty data traffic, a higher weight may be assigned to the rate of change of the data rate.

The algorithm operating on the BS, in one embodiment, is implemented in software. The BS executes the software which, by taking into consideration the three above mentioned factors, maximizes the throughput of the base station as well as the data rate of individual subscriber units, and attempts to maximize the quality of service, or grade of service, of data packet transmissions anticipated or negotiated by or on behalf of the subscriber units. Software also executes on each of the subscriber units to provide the

information relating to the three factors on which the base station decides to allocate RF channels among the subscriber units.

With reference to figure 1, the process of RF channel allocation, in accordance with an embodiment of the present invention, is as follows. The process starts at 105

5 with both BS and a number of subscriber units powered up and ready for operation. An individual SU, say SU1, monitors the size of a data transmission queue on the SU, and the rate of change in the data rate (which can be derived from the rate of change in the size of the data transmission queue on the SU). Based on these factors, the SU determines whether it needs additional or less RF channels to transmit data in the queue
10 waiting to be transmitted to the BS. In one embodiment, the SU further monitors the number of RF channels already allocated to the SU, if any, and considers this allocation as well in determining whether it needs additional RF channels to transmit data waiting in the data transmission queue to the BS.

There is also a notion of a moving average data transmission queue length. When
15 the SU's data transmission buffer reaches a threshold, say 80% full (wherein such percentage may be based on the instantaneous and/or moving average buffer size), the SU transmits an RF channel allocation request at 110, requesting a number of RF channels to be allocated to the SU.

It should be noted that while the above discussion refers to a threshold for the size
20 of the data transmission buffer for a subscriber unit, a threshold may be associated with any one or more of the factors being given consideration under the algorithm contemplated by the present invention. When any one threshold is reached or exceeded for some minimum amount or period of time, a request to allocate an additional RF

channel resource is generated by the SU and transmitted to the BS. In essence, then, the thresholds determine how many RF channels are allocated per subscriber unit. Moreover, different thresholds can be set depending on the level or quality of service desired for any one particular subscriber unit.

5 In one embodiment of the present invention, the requesting subscriber unit may alter or adjust the threshold at which a request for additional allocation of RF channel resources is issued based on the number of channels already allocated to the subscriber unit. Essentially, then, the number of channels already allocated to the subscriber unit provides the subscriber unit with the ability to generate requests of varying levels of
10 priority. The more channels already allocated to the subscriber unit, the higher the threshold before triggering a request for allocation of yet another RF channel. In one embodiment, a subscriber unit may request and receive up to four simultaneous RF channels. It is appreciated, however, that the number of channels allocated may differ more or less than four without departing from the present invention.

15 At the BS, the RF channel allocation algorithm receives the RF channel allocation request from SU1, and at 115 examines whether there are available RF channels that can be allocated to SU1. The BS monitors the RF channel usage of all subscriber units within the domain of the BS and the maximum RF channel capacity of the system. The BS therefore is able to determine whether any RF channels are immediately available for
20 allocation to SU1. The BS further monitors the number of RF channels allocated per subscriber unit, including the SU from which it received the RF channel allocation request. The BS also monitors outstanding requests for RF channels from other

subscriber units, and the length of time an RF channel is utilized by a particular subscriber unit.

In one embodiment, the BS grants the request for RF channel allocation if the BS detects available channels, at step 120. The BS, for example, maintains a queue of available channels, and allocates one of the available RF channels in the queue to the requesting subscriber unit.

If there are more requests for RF channel allocation than there are RF channels, the BS decides whether to honor the RF channel allocation request from a particular subscriber unit, or deny the request. The BS station, in making such a decision, considers three factors, not necessarily in the order as discussed as follows. The BS examines, at 125, the instantaneous quantity or amount of data queued at the SU for transmission to the BS. This information typically is provided to the BS in the allocation request transmitted from the SU to the BS. The BS further examines the rate of change in the amount of data queued at the SU for transmission, at 130. The rate of change can be calculated and monitored on the BS based on previous requests for resources transmitted by the same SU, or such information may be calculated, monitored, and provided by the SU in the resource allocation request. It is appreciated that as the data transmission queue on the SU increases in length, or the rate of change in the amount of data queued for transmission from the SU increases, the resource allocation request from the SU may carry greater weight with the BS. At 135, the BS additionally examines the amount of time that the SU has been allocated one or more RF channels. In one embodiment, the BS associates a weight to one or more of the factors and considers at 140 whether to

honor the request in order to provide a certain quality or grade of service, based on the three factors and any weights associated therewith.

If the desired quality of service is not being met for the requesting subscriber unit, the BS allocates an RF channel resource to the requesting subscriber unit. In one

5 embodiment, as RF channels no longer are utilized by other of the subscriber units, a queue is maintained on the BS, the queue tracking RF channels that are no longer in use and can be torn down for later setup and use by other subscriber units. In allocating an RF channel to the requesting subscriber unit, the BS first checks whether the teardown queue has one or more RF channels that can be torn down and then set up for the

10 requesting subscriber unit. If there are no RF channels to be torn down, the BS, at 145, having knowledge of all RF channels in use, and the number of RF channels allocated to each subscriber unit in the domain of the BS, selects an RF channel to deallocate from a subscriber unit. At 150, the now available RF channel is allocated to the requesting subscriber unit. In one embodiment, the BS will tear down an RF channel allocated to a
15 subscriber unit with the most number of RF channels allocated for the longest time.

Alternatively, the BS maintains track of the utilization of each RF channel resource and tears down/deallocates a least used RF channel. It is appreciated that the number and duration of channel allocations per subscriber unit may be maintained in a separate queue.

20 If, at 140, the grade of service desired for a subscriber unit is being met or exceeded, given that no RF channel resources are available, the BS does not honor the request for RF channel allocation and waits to receive another RF channel resource request at 110.

All requests received by the BS from the subscriber units are maintained in a queue to both track and maintain the order of requests as they arrive at the BS. If there are entries in the tear down queue, that is, the queue that maintains the RF channels no longer utilized by a subscriber unit, the first, or oldest, channel allocation request may be honored by granting thereto the first, or oldest, RF channel tracked in the tear down queue. If there are no entries in the tear down queue, the BS will add a subscriber unit's request for RF channel allocation to a set up queue in the order it was received.

In one embodiment, the RF channel allocation requests from a particular subscriber unit may be prioritized in relation to the RF channel allocation requests from other subscriber units, based on the number of channels already allocated to the particular subscriber unit, and/or the number of RF channels being requested by the particular subscriber unit.

Thus far, the discussion of the present invention has focused on the SU requesting an RF channel to transmit information to the BS in the uplink direction, that is, from the SU to the BS. It is appreciated, however, that the invention operates essentially in the same manner when the BS is requesting an RF channel to transmit information to the SU in the downlink direction, that is, from the BS to the SU. The significant distinctions in allocating RF channels for downlink versus uplink transmissions, of course, is that the BS becomes the device that initiates the resource allocation request, and is the device that monitors all the factors considered by the algorithm. Thus, for example, the BS monitors the depth of the data transmission queue at the BS for a particular SU and the rate of change in the depth of the data transmission queue, rather than the SU monitoring such factors. The BS further monitors the number of RF channels already allocated between

the BS and the SU to which it has data to transmit, and the amount of time the already allocated RF channels have been utilized for transmitting data from the BS to the SU.

After considering all these factors, the BS makes a decision to initiate a resource allocation request, in particular, a RF channel allocation request. The request is internal,

- 5 that is, the “bandwidth on demand” software executing on the BS causes the BS to initiate an internal request to itself to allocate a channel for communicating data in a data transmission queue in the BS with the SU.

The present invention thus attempts to grant a new subscriber unit an RF channel, while minimizing latency and maximizing throughput in the circuit switched network.

Attorney Docket No. 15685.P074